ISO/IEC JTC 1/SC 29/WG 04 N0341

ISO/IEC JTC 1/SC 29/WG 04

MPEG Video Coding

Convenorship: CN

Document type:	Output Document
Title:	Verification test report of MPEG immersive video
Status:	Approved
Date of document:	2023-06-02
Source:	ISO/IEC JTC 1/SC 29/WG 04
Expected action:	None
Action due date:	None
No. of pages:	12 (without cover page)
Email of Convenor:	yul@zju.edu.cn
Committee URL:	https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-4

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 04 MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 04 N 0341
April 2023, Antalya

Title	**Verification test report of MPEG immersive video**
Source	**WG 04, MPEG Video Coding**
Status	**Approved**
Serial Number	**22688**

Abstract

This document provides the test report of the MPEG immersive video (MIV) verification test, demonstrating the performance of the new standard in comparison with the previous state-of-the-art MPEG video standard for coding multiple views – the multi-view extension of HEVC (MV-HEVC). A formal subjective quality evaluation with "naïve" test subjects watching pre-defined pose traces in an immersive scene was performed. According to the results, the average mean opinion score (MOS) savings of the MIV Main profile range from 1.20 to 4.69 (in 11-grade quality scale) for tested sequences, while for the MIV Geometry Absent profile the average savings was equal to 2.94.

1. Introduction

The MIV standard is part of ISO/IEC 23090 MPEG-I, a collection of standards to digitally represent immersive media. This standard features the compression of immersive video content, also known as volumetric video, in which a real or virtual 3D scene is captured by multiple real or virtual cameras. It enables storage and distribution of immersive video content over existing and future networks for playback with 6 degrees of freedom (6 DoF) of view position and orientation within a limited viewing space and with different fields of view depending on the capture setup. The basis of the MIV encoder's operation is to perform preliminary processing of a multi-view sequence and generate metadata related to the transmitted views. Then, the sequence in the form of atlases (containing input views or their fragments) is encoded using any method of video compressing, for example, using an encoder compliant with any MPEG encoding standard.

The MIV standard offers a range of profiles that can be adjusted to suit various levels of bandwidth or decoding capabilities on the client's end. Alongside the MIV Main profile [1], which relies on geometry and includes embedded occupancy and texture (often referred to as MVD for multiple video + depth), there is an MIV Extended profile that potentially separates occupancy from geometry and enables the use of a transparency attribute. This profile also includes a Restricted Geometry sub-profile that facilitates MPI format delivery [2]. Lastly, the MIV Geometry Absent profile [3] allows the client to generate geometry information locally or in the cloud.

This document reports the MPEG immersive video verification test, prepared to demonstrate the performance of the new standard in comparison with the previous state-of-the-art coding of immersive video. The configurations used for generating tested videos of MV-HEVC [4] (the multi-view extension of HEVC, selected as the anchor codec) and tested MIV profiles (MIV Main and MIV Geometry Absent) are presented in Section 2, while the test setup and its results presented in Sections 3 and 4. Section 5 presents a discussion on acquired results and concludes the report.

The report summarizes information presented in “Plan for verification test of MPEG immersive video” [5] by WG4 and “Results of visual dry run testing for MIV verification test” [6] provided by AG5. Both documents are non-public, therefore, relevant parts were reproduced in this test report in order to be shared with organizations interested in MIV.

2. Tested configurations and test material

2.1. Summary of tested configurations

The data sets were compressed with the anchor software and the MIV test model software TMIV with target bitrates 5, 17, 28, and 40 Mbit/s.

- Best reference

The best reference serves as an indicator of the quality produced if no compression is applied to the data set before rendering. Accordingly, the video sequences for the best reference were generated from all views and depth maps using TMIV renderer (VWS) without any pruning of input views or compression.

- MV-HEVC (Anchor)

The encoding was performed with the MV-HEVC reference software HTM 13.0, and the rendering using the Reference View Synthesizer (RVS) 4.0 synthesizer for all data sets and all rate points.

- MIV Main profile

The encoding was performed with TMIV v14.0 VT branch, MIV Main profile, using the VWS synthesizer. Bitstreams were generated using HEVC for compression of the reported data sets and rate points.

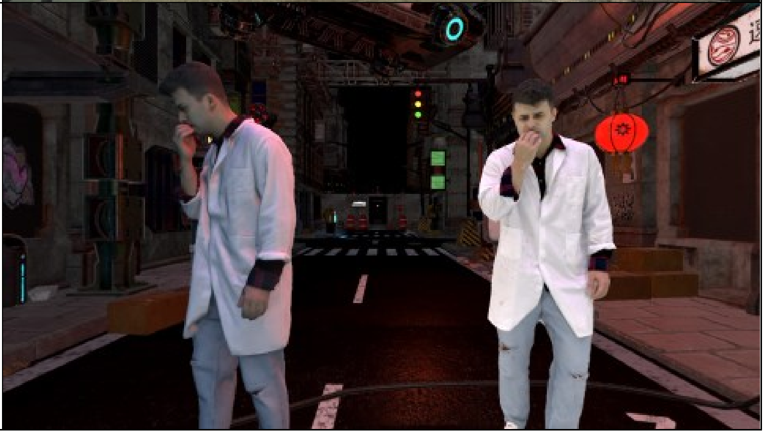
- MIV Geometry Absent profile

The encoding was performed with TMIV v14.0 VT branch, MIV Geometry Sbsent profile, using the VWS synthesizer. Bitstreams were generated using HEVC for compression. For this configuration, the natural content data sets were encoded at all rate points.

The detailed configuration and its description is provided in the following subsections.

2.2. Test material

Following set of four multiview sequences was used for the verification test purposes. The table below provides names of used data sets and an example of still frame from best reference pose trace. Characteristics and descriptions of these sequences can be found in Annex A of common test conditions for MPEG immersive video [7].

Sequence	An example of frame from best reference pose trace
F (Guitarist)	
S (CBABasketball)	
W (Dancing)	
X (Cyberpunk)	

2.3. MV-HEVC anchor generation

- The number of basic views is driven by the pixel rate constraint. The selection of the basic views is simple, not based on camera parameters or analysis of the pixels, to reflect the status before the start of the MIV project.
- Configuration files for Reference View Synthesizer (RVS) 4.0, used for synthesizing views, were provided by ULB and Philips. RVS was updated by InterDigital to have the functionality of rendering pose traces similar to one included in TMIV.
- The following process was used for generating the MV-HEVC anchor:

1. Use basic view selector.
2. Use HDRTools0.18 to convert depth maps of the selected views (obtained in the first step) from yuv420p16le to yuv420p10le. A sample config file to use but proponents need to adjust the parameters according to the sequences is available with this document in configuration directory (*HDRConvert_yuv420p16leToyuv420p10le.cfg)*
3. Get HTM13.0 (VC10 required to compile and build or you can try migrating to newer versions). Build with macro HEVC_EXT set to 1 in TypeDef.h to generate the MV-HEVC executable.
4. Run MV-HEVC with the attached config files for both the texture and depth content (proponents need to adjust the parameters according to the sequences). Please note that QPs and paths in the configuration files need to be updated.
5. Run RVS 4.0 such that it inputs the decoded MV-HEVC views and reconstructs the other missing views (proponents need to adjust the provided configuration according to the sequences). Note: *BitDepthDepth values in sequence parameters JSONs have to be changed to 10 from 16*.

Texture QP values to achieve a chosen set of bit rates of 5, 17, 28, and 40 Mbps:

Sequence	QP1 (40 Mbps)	QP2 (28 Mbps)	QP3 (17 Mbps)	QP4 (5 Mbps)
F	16	19	23	32
W	13	17	20	25
X	12	16	20	25
S	22	25	29	39

Depth QP values were decreased by 10 (QPd = QPt – 10). This calculation was kept simple to reflect the status of the state of the art before the start of the MIV project.

Prerequisites / software versions / configuration:

Software	**Version**	**Link**	**Configuration**
MV-HEVC (HTM)	13.0	HTM-13.0	Available with this document in configuration directory: - EE2.1VT_MVHEVC_D_geo.cfg - EE2.1VT_MVHEVC_D_tex.cfg
HDRTools	0.18	HDRTools/tree/v0.18	Available with this document in configuration directory: - HDRConvert_yuv420p16leToyuv420p10le.cfg

Software	Version	Link	Configuration
Basic views selector	-	vt_anchor_select_cameras.py	Type: python vt_anchor_select_cameras.py --help
RVS	4.0	rvs/-/tree/v4.0	Default configuration from rvs/-/tree/v4.0/config_files Pose traces files (.csv) are available in: tmiv/-/tree/v14.1/config/vt/pose_traces. Used pose traces: X – p03, W – p04, S – p04, F – p02 Parameters for sequences are available in: tmiv/-/tree/v14.1/config/vt/sequences

2.4. Generation of pose traces for tested MIV profiles

2.4.1. MIV Main

Texture QP values used in HM to achieve a chosen set of bit rates of 5, 17, 28, and 40 Mbps:

Sequence	QP1 (40 Mbps)	QP2 (28 Mbps)	QP3 (17 Mbps)	QP4 (5 Mbps)
F	21	25	30	41
W	25	28	33	44
X	21	24	29	42
S	21	25	30	41

Depth QP values were calculated using the equation provided in common test conditions for MPEG immersive video [7].

Prerequisites / software versions / configuration

Software	Version	Link	Configuration
TMIV	14.1	tmiv/-/tree/v14.1	tmiv/-/tree/v14.1/config/vt Pose traces files (.csv) are available in: tmiv/-/tree/v14.1/config/vt/pose_traces. Used pose traces: X – p03, W – p04, S – p04, F – p02 Parameters for sequences are available in: tmiv/-/tree/v14.1/config/vt/sequences
HM	16.16	HM.git	tmiv/-/tree/v7.0/ctc_config/miv_anchor (HM cfg for TMIV 7.0)

A Python script is provided in tmiv/-/tree/v14.1/scripts/test and is able to run the configuration used in MIV Main verification test: use **vt1_vt1: Encode VT MIV main, decode VT MIV main.**

2.4.2. MIV Geometry Absent

Texture QP values used in HM to achieve a chosen set of bit rates of 5, 17, 28, and 40 Mbps:

Sequence	QP1 (40 Mbps)	QP2 (28 Mbps)	QP3 (17 Mbps)	QP4 (5 Mbps)
S	20	24	27	31

Prerequisites / software versions / configuration

Software	**Version**	**Link**	**Configuration**
TMIV	14.1	tmiv/-/tree/v14.1	tmiv/-/tree/v14.1/config/ctc/miv_dsde_anchor Pose traces files (.csv) are available in: tmiv/-/tree/v14.1/config/vt/pose_traces. Used pose trace: S – p04 Parameters for sequences are available in: tmiv/-/tree/v14.1/config/vt/sequences
HM	16.16	HM.git	tmiv/-/tree/v7.0/ctc_config/miv_anchor (HM cfg for TMIV 7.0)
IVDE	7.0	ivde/-/tree/v7.0	As in ivde/-/blob/v7.0/README.md with following changes: "TotalNumberOfFrames": 97 “NumOfThreads”: 1 “FeaturesSkipThresh”: 0.25

3. Test setup

3.1. Logistics

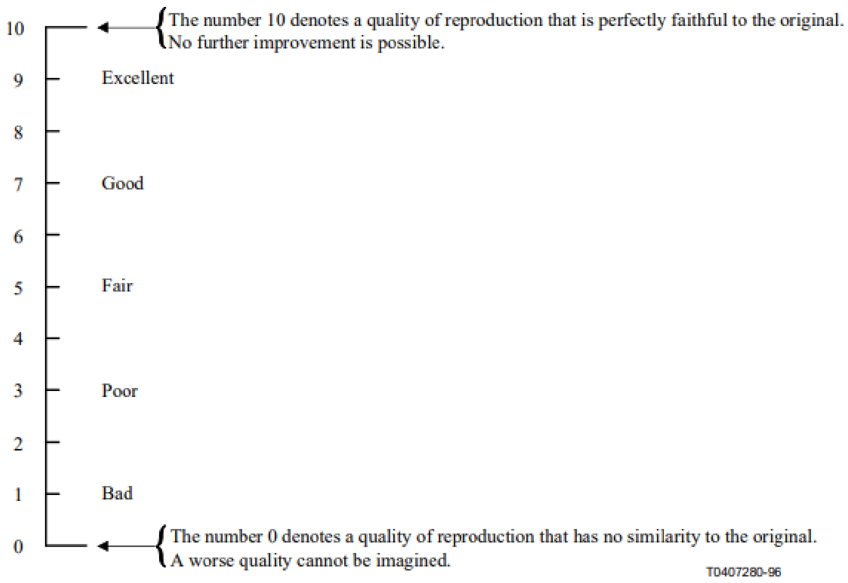
Test Site	**RWTH Aachen University**
Display, size (resolution setting), connection	- 1× Sony 55” PVM X550 (3840×2160), Quad-SDI - 1× LG OLED65CX (3840×2160), HDMI - 2× LG OLED55G19LA (3840×2160), HDMI The X550 was driven by a PC with a DeckLink Extreme 4G video board via Quad-SDI, the signal further converted to HDMI by an AJA Hi5-4K-Plus converter and sent in parallel to the three LG displays via an HDMI splitter.
Viewing distance	2 viewers at 2H per display. In some sessions, 3 viewers were placed in front of the LG OLED65CX.
Viewing angle	70° for 2 viewers, 90° for the central viewer in case of three viewers
Total number of viewers	30 (1 female, 29 males; age 18-24, 10 different nationalities), all screened for visual acuity and normal colour vision

The viewers were acquired among students at RWTH Aachen university and from a local school.

3.2. Test method

The ACR-HR (Absolute Category Rating – Hidden Reference) methodology as described in recommendation ITU-T P.910 **Error! Reference source not found.** was used in the viewing session. ACR is a quality judgment where the processed video sequences (PVS) are presented one at a time and are rated independently from each other, on a quality scale. After each presentation the subjects are asked to evaluate the quality of the sequence shown.

The presentation time was 9.7 seconds (291 frames) for all PVS, while the voting time was 5 seconds. The 11-grade quality scale of ITU-T P.910 B.2 (presented below) has been used. The test procedure includes a reference version of each test sequence shown as any other test stimulus.


10
9
8
7
6
5
4
3
2
1
0
The number 10 denotes a quality of reproduction that is perfectly faithful to the original.
No further improvement is possible.
Excellent
Good
Fair
Poor
Bad
The number 0 denotes a quality of reproduction that has no similarity to the original.
A worse quality cannot be imagined.
T0407280-96

3.3. Test design

A total of 81 rendered video sequences were provided for visual evaluation. The video sequences were presented in four test sessions of 32 PVS each. The duration of each session was about 8.2 s. The test sessions included a calibration phase in the first session and stabilization phases plus trapping sequences in each session. The trapping sequences included best reference sequences and sequences considered to be rated very low, which were repeatedly presented over the test sessions. The data of the calibration phase and the stabilization phase were not regarded in the evaluation.

Before the test, the basic concept of MIV and the process of generating the tested video sequences by pre-defined pose traces in an immersive scene was explained to the viewers. In the training session, the grading scale was explained to the viewers. They were familiarized with the session layout and the scoring sheets. The best reference and a set of example sequences from the test set were presented and the viewers were advised to find their personal scoring for the presented sequences. The viewers were advised to stay totally silent during the sessions and to only express their opinion on the scoring sheet.

4. Test results and analysis

Results shown below were acquired by AG 05 and are available in [5], which describes full methodology of formal subjective evaluation according to the Absolute Category Rating (ACR) method of ITU-T P.910 with naïve test subjects.

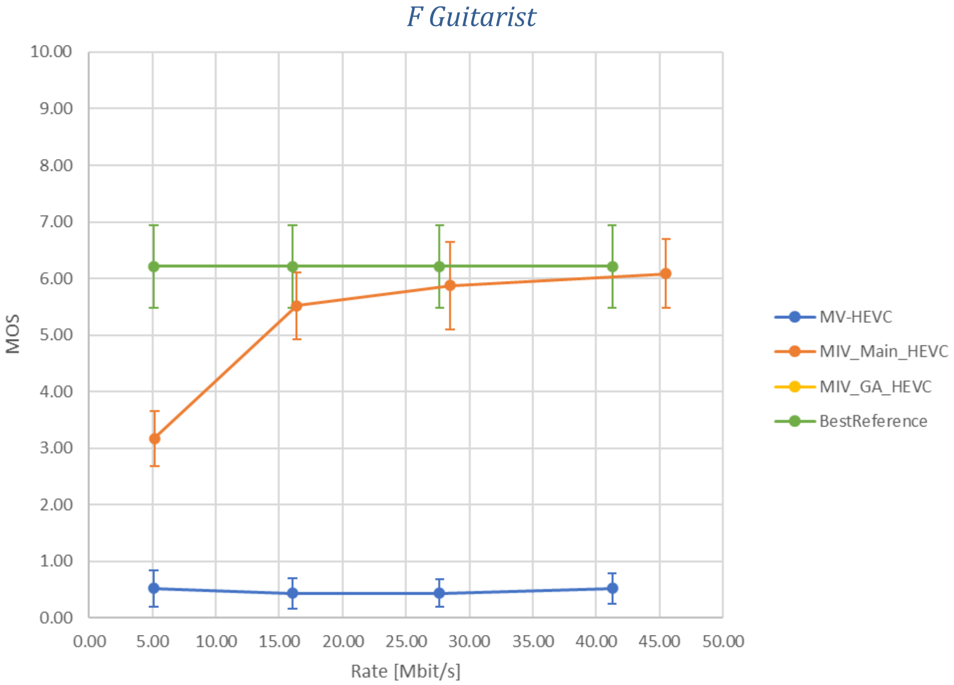
4.1. Data processing

As a first step, the received scores of the viewers were screened for the scoring of the ‘trapping’ best reference sequences (same sequence across sessions). Candidates with a divergence of more than 2 steps on the grading scale were excluded from the subsequence evaluation. This applied to 6 viewers. The screening of the ‘trapping’ low-quality sequences did not lead to discarding any viewers.

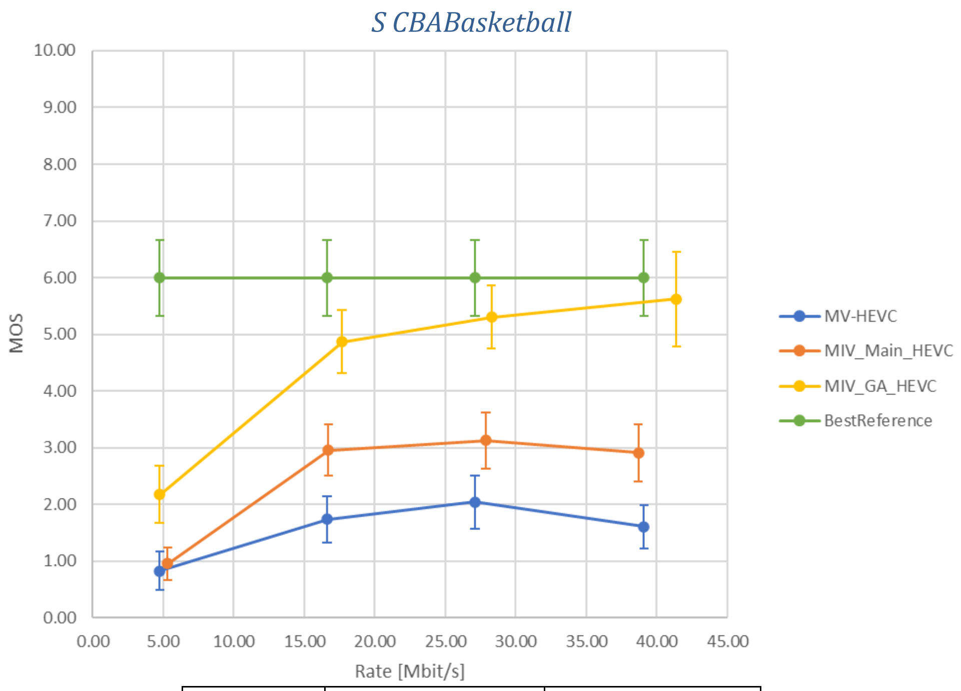
As a next step, the scores were screened for their z-scores. All scores were found to be below a z-score of 3. The scores were further screened for the Pearson correlation coefficient (PCC) for each session. The scores of viewers with a PCC lower than 0.75 were excluded in 3 cases.

4.2. Results comparing TMIV to the MV-HEVC anchor

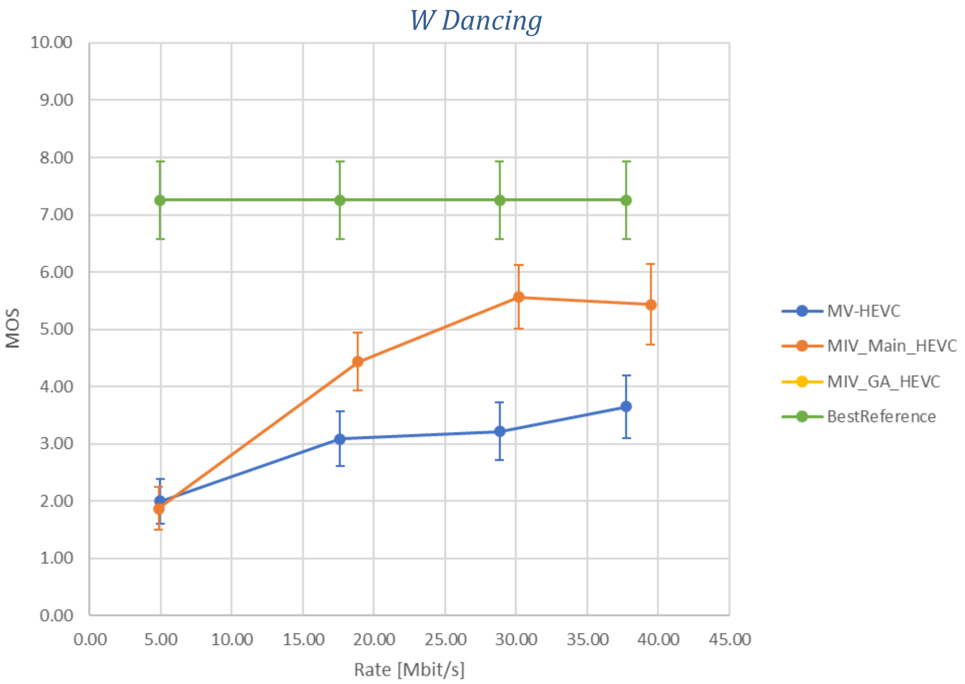
In the following subsections, the MOS-over-rate plots for the test data sets are shown. Additionally, tables are included which report the delta MOS between the anchor (MV-HEVC) and the scheme under investigation (MIV Main profile or MIV GA profile). The delta MOS values indicate the expected quality improvement in terms of MOS when MIV is used instead of the anchor scheme. They are only computed if the confidence intervals of the two schemes under comparison are not overlapping. Otherwise, the table entry is marked with "n/a".



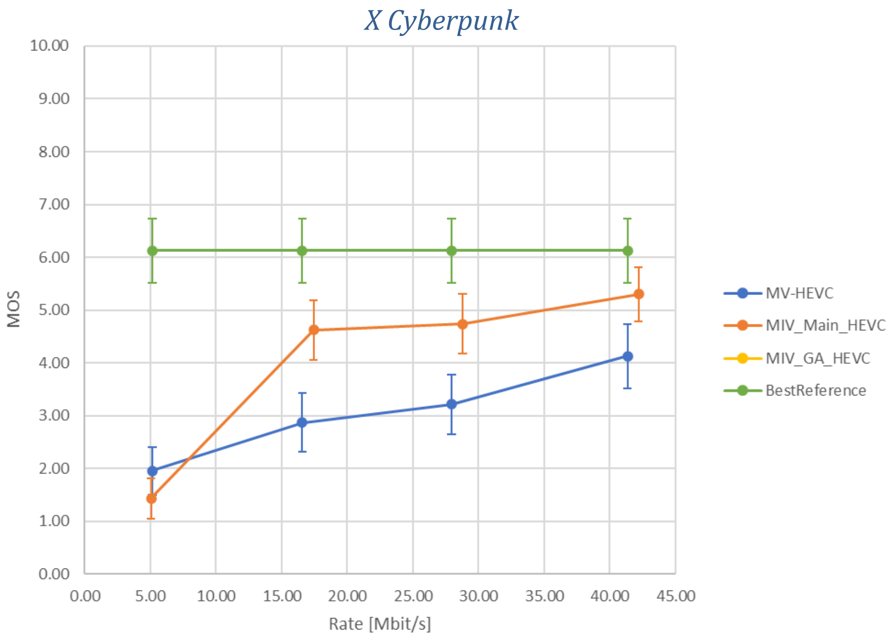
DeltaMOS	MIV_Main_HEVC
QP1	5.57
QP2	5.44
QP3	5.09
QP4	2.65



DeltaMOS	MIV_Main_HEVC	MIV_GA_HEVC
QP1	1.30	4.02
QP2	1.09	3.26
QP3	1.22	3.13
QP4	n/a	1.35



DeltaMOS	MIV_Main_HEVC
QP1	1.78
QP2	2.35
QP3	1.35
QP4	n/a



DeltaMOS	MIV_Main_HEVC
QP1	1.17
QP2	1.52
QP3	1.76
QP4	n/a

5. Discussion and conclusions

The MOS results generally demonstrate a clear benefit of MIV over the anchor MV-HEVC. Due to the partially non-monotonic characteristics of the curves, a computation of BD MOS savings has not been performed. Instead, the delta MOS between the anchor and MIV are computed. According to the corresponding results the average delta MOS savings for the MIV Main profile range from 1.20 to 4.69.

For the data set F (Guitarist) specifically, the compression scheme of the anchor is not capable of providing a result suitable for the application. It must be noted that, at the same time, the MIV scheme already reaches saturation at the MOS of the best reference (i.e., overlapping confidence intervals with the best reference) at the second lowest rate point. This indicates the compression efficiency to be up to a degree where degradations may be attributed dominantly (or only) to the rendering.

For MIV Main profile, overlapping with the best reference is observed once more for the highest rate point of the X (Cyberpunk) sequence. For the S (CBABasketball) data set, a significant improvement over the anchor as well as the MIV Main profile is observed for the MIV GA profile. The average delta MOS over the rate points is 2.94 in this case, with three out of four rate points have overlapping confidence intervals with the best reference, indicating a tendency towards irrelevance of the compression impact to the assessed video sequence. Generally, the plots reveal that some saturation effects seem to occur at higher rate points for most data sets. This saturation may be observed at a lower MOS than reported for the best reference. Results have shown also some limitations of the available source material, as some reference pose traces that did not include any visible rendering artefacts were scored relatively low (e.g. X – Cyberpunk data set), as viewers were not aware of the quality of source material.

It is noted that the reported test treats the rendered MIV datasets as conventional 2D video which may imply a more critical scoring than to be expected in an interactive application. The test method, being widely known and established, was deliberately chosen for its significant discriminatory power. Further, while some MOS scores can be perceived as relatively low in this test, the methodology of tests does not reflect the possibility of free virtual navigation for the viewer. Therefore, the presented results represent the quality perceived by viewers without awareness of the full possibilities of the presented technology. Nevertheless, having in mind providing a safe, fully reproducible test, it was required to present pre-rendered videos to viewers during the verification test.

To present the full capabilities of the tested standard, the attendees of the 142nd MPEG meeting had an opportunity to watch a presentation of real-time MIV decoding and high-quality rendering of virtual views performed on a consumer-grade smartphone [9] or on a laptop [10]. In the first demonstration, a menu structure of the presented application allowed for navigation to select clips to play. The available content comprised two MVD sequences, two MPI-based sequences (encoded using relevant MIV profiles), and six point-cloud-based sequences (encoding using V-PCC). 6DoF interaction was possible using touch screen control or a gyroscope. In the latter demonstration, the interaction with two presented sequences was possible through the face tracker, capturing video of the viewer using the built-in camera. It enabled continuous synthesis of new views which corresponded to the position of the user. The positive reception of these demonstrations [11] clearly shows evidence of sufficient maturity of MIV standard, ready to be successfully used in practical applications.

6. References

[1] J.M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V.K.M. Vadakital, L. Yu, “MPEG Immersive Video Coding Standard,” Proceedings of the IEEE, vol. 109, no. 9, pp. 1521-1536, 2021.

[2] V.K.M. Vadakital, A. Dziembowski, G. Lafruit, F. Thudor, G. Lee and P.R. Alface, “The MPEG Immersive Video Standard—Current Status and Future Outlook,” IEEE MultiMedia, vol. 29, no. 3, pp. 101-111, 2022.

[3] D. Mieloch, P. Garus, M. Milovanović, J. Jung, J. Y. Jeong, S. L. Ravi, B. Salahieh, “Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6360-6374, Sept. 2022.

[4] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, “Overview of the Multi-view and 3D Extensions of High Efficiency Video Coding”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, Issue 1, pp. 35-49, Sept. 2015.

[5] ISO/IEC JCT 1/SC 29/WG 4, “Verification test preparations of MPEG immersive video,” Doc. N0235, July 2022.

[6] ISO/IEC JCT 1/SC 29/AG 5, “Results of dryrun for MIV verification test,” Doc. N0082, January 2023.

[7] ISO/IEC JCT 1/SC 29/WG 4, “Common test conditions for MPEG immersive video,” Doc. N0307, January 2023.

[8] Recommendation ITU-T P.910 (2008), “Subjective video quality assessment methods for multimedia applications”.

[9] B. Kroon, C. Guede, P. Fontaine, B. Sonneveldt, C. Varekamp, “Real-time decoding and rendering demo on a smart phone,” ISO/IEC JCT 1/SC 29/WG 4, Doc. M63058, April 2023, Antalya.

[10] G. Lee, H.-C. Shin, J. Y. Jeong, K.-J. Oh, W.-S. Cheong, “Prototype MIV player for demonstration,” ISO/IEC JCT 1/SC 29/WG 4, Doc. M63219, April 2023, Antalya.

[11] ISO/IEC JCT 1/SC 29/WG 4, “Report of 11th meeting,” Doc. N0321, April 2023, Antalya.